PSYCHOMETRIKA—VOL. 57. NO. 3, 357–369
SEPTEMBER 1992

CONSTRUCTING A COVARIANCE MATRIX THAT YIELDS A SPECIFIED MINIMIZER AND A SPECIFIED MINIMUM DISCREPANCY FUNCTION VALUE

ROBERT CUDECK

UNIVERSITY OF MINNESOTA

MICHAEL W. BROWNE

OHIO STATE UNIVERSITY

A method is presented for constructing a covariance matrix Σ_0^* that is the sum of a matrix $\Sigma(\gamma_0)$ that satisfies a specified model and a perturbation matrix, $\mathbf{E}$, such that $\Sigma_0^* = \Sigma(\gamma_0) + \mathbf{E}$. The perturbation matrix is chosen in such a manner that a class of discrepancy functions $F(\Sigma_0^*, \Sigma(\gamma_0))$, which includes normal theory maximum likelihood as a special case, has the prespecified parameter value γ_0 as minimizer and a prespecified minimum δ. A matrix constructed in this way seems particularly valuable for Monte Carlo experiments as the covariance matrix for a population in which the model does not hold exactly. This may be a more realistic conceptualization in many instances. An example is presented in which this procedure is employed to generate a covariance matrix among nonnormal, ordered categorical variables which is then used to study the performance of a factor analysis estimator.

Keywords: Monte Carlo experiments, covariance structure analysis, factor analysis, model misspecification.

1. Introduction

Psychometric methods are often developed assuming that the population covariance matrix has a particular structure. Estimators of a model or tests of model fit are generally derived under the hypothesis that these structural assumptions are true, at which point two practical questions often arise. First, when the model and all other assumptions hold, how accurate is a proposed estimator or test statistic under ideal conditions? Second, if the assumptions are violated, in what ways is the performance of the estimator or testing procedure affected? If analytic methods cannot be used to address these questions, then Monte Carlo experiments can be designed to investigate them instead. The most informative Monte Carlo experiments reflect those characteristics of the population, sampling process, and anticipated data that are known or presumed to be most reasonable.

In the factor analysis literature a variety of Monte Carlo studies have been conducted to investigate various features of estimators of the unrestricted or restricted models. These experiments have most often been based on a population correlation matrix that satisfies a simple factor analysis model exactly. While undeniably useful, these experiments can be criticized on the grounds that it is unlikely that any simple factor analysis model perfectly fits a population matrix.

We are grateful to Alexander Shapiro for suggesting the proof of the optimality of the solution in section 2.

Requests for reprints or a copy of the computer program that implements this method should be sent to Robert Cudeck, Department of Psychology, University of Minnesota, 75 East River Road, Minneapolis, MN 55455, or to Michael W. Browne, Departments of Psychology and Statistics, Ohio State University, Columbus, OH 43210.

0033-3123/92/0900-91312.75/0$00.75/0

In a notable departure from the practice of assuming a factor analysis model holds exactly, Tucker, Koopman, and Linn (1969) argued that actual data do not appear to follow such simple models, even in large representative samples with high quality data. Consequently, the generality of findings from a Monte Carlo study that assumes a particular model holds can be questioned. As an alternative, Tucker et al. suggested that more realistic population matrices be simulated by including three kinds of factors. Major factors are a small group of dominant latent variables, most often with a simple pattern. Unique factors are the traditional specific effect plus error associated with each variable, all mutually uncorrelated. Minor factors, in practice fifty or even more in number, represent small sources of covariance among the variables. Matrices generated by this method seem more in keeping with actual data in the sense that although a small number of major factors is present, no simple factor analysis model holds exactly. In substantive terms, the influences represented by the minor factors reflect any of a variety of unsystematic or unknown aspects of the process that generates the data, such as non-random error or localized method effects. The idea in general is that the population covariance matrix is made up of a particular structure, plus additional elements of covariance representing lack of fit. The latter are not typically considered further. This method has been effectively used in several empirical studies (e.g., Hakstian, Rogers, & Cattell, 1982; Laughlin, 1979).

The purpose of this article is to describe another method for constructing a covariance matrix in which a particular structure holds approximately. The specified departure or lack of fit between the population matrix and the model is operationalized as an *exact value* of a discrepancy function. The primary structure need not be a factor analysis model. Any model in the general class of covariance structures is appropriate. A simple modification to the procedure allows structured mean vectors to be included as well. Therefore, the method may be considered in a variety of applied problems that have been studied by Monte Carlo methods, including analysis of variance or regression models with an assumed covariance structure, and certain multidimensional scaling models that use correlations or covariances as proximity measures. This work was motivated by Shapiro and Browne's (1988) suggestion that the sequence of systematic error terms in a population drift process could be chosen so as to avoid giving the impression that estimators are asymptotically biased.

2. The Basic Procedure

Let $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ denote a $(p \times p)$ symmetric matrix-valued function of a parameter vector $\boldsymbol{\gamma}' = (\gamma_1, \ldots, \gamma_q)$, where $q \leq \frac{1}{2}p(p+1)$, and let $\boldsymbol{\gamma}_0$ be a particular value within the admissible region $\boldsymbol{\Gamma}$ at which $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$ is positive-definite. We assume that the elements of the model matrix, $\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \{\sigma_{jk}(\boldsymbol{\gamma})\}$, are continuously differentiable functions in a neighborhood of $\boldsymbol{\gamma}_0$, and assume also that $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ is identified at the point $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$. Let $\mathbf{E}$ $(p \times p)$ be a symmetric matrix such that the sum,

$$\boldsymbol{\Sigma}_0^* = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0) + \mathbf{E}, \tag{1}$$

is positive-definite. We consider discrepancy functions of the form

$$F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma})) = \tfrac{1}{2} \operatorname{tr} [\mathbf{W}^{-1}(\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma}))]^2, \tag{2}$$

where $\mathbf{W}$ is a fixed matrix that does not depend on $\mathbf{E}$. Two versions in particular are of interest. The first case corresponds to ordinary least squares, which is obtained when in (2) the substitution $\mathbf{W} = \mathbf{I}$ is made. The second case is normal theory maximum likelihood. Let $\boldsymbol{\gamma}_{\mathrm{ML}}$ be the minimizer of

$$M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma})) = \ln |\boldsymbol{\Sigma}(\boldsymbol{\gamma})| - \ln |\boldsymbol{\Sigma}_0^*| + \operatorname{tr} [\boldsymbol{\Sigma}_0^* \boldsymbol{\Sigma}(\boldsymbol{\gamma})^{-1}] - p. \tag{3}$$

If $\mathbf{W} = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_{ML})$, the minimizer of (2) is the same as the minimizer of the maximum likelihood discrepancy function (3), although the respective minima are not in general equal (Browne, 1974).

The problem to be addressed can be stated as follows. Given a particular value $\boldsymbol{\gamma}_0$ for the parameter vector and a value $\delta > 0$ for the lack of fit, we seek a matrix $\mathbf{E}$ in (1) such that

(i) the minimizer of $F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))$ is the required value $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0$, and
(ii) at the point $\boldsymbol{\gamma}_0$, the minimum function value will be one of the following:

$$M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) = \delta, \qquad \text{if } \mathbf{W} = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0),$$

$$F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) = \delta, \qquad \text{otherwise,}$$

where δ is a prespecified value. A solution can be developed by considering the gradient of (2), and exploiting the relationship between $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$ and $\boldsymbol{\Sigma}_0^*$. Let $\dot{\boldsymbol{\Sigma}}_i = [\partial \boldsymbol{\Sigma}(\boldsymbol{\gamma})/\partial \gamma_i]$ be the derivative of the model with respect to the i-th element of $\boldsymbol{\gamma}$. It can be shown that the derivative of F with respect to γ_i is

$$\frac{\partial F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))}{\partial \gamma_i} = -\operatorname{tr} [(\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma}))\mathbf{W}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{W}^{-1}], \tag{4}$$

(for example, Browne, 1974, sec. 3). With the substitution $\mathbf{W} = \boldsymbol{\Sigma}(\boldsymbol{\gamma})$, (4) is also the derivative of M. Because of this equality, the following results, which are stated generally in terms of the derivative of F, also apply more specifically to the derivative of M.

Let $\mathbf{A}$ and $\mathbf{B}$ be symmetric matrices of order p. Let vecs $(\mathbf{A}) = (a_{11}, a_{12}, a_{22}, a_{13}, \ldots, a_{pp})'$ denote the vector of length $p^* = \frac{1}{2}(p^2 + p)$ obtained by stacking the nonduplicated elements of $\mathbf{A}$, including the diagonal, and let vec $(\mathbf{A}) = (a_{11}, a_{21}, \ldots, a_{p1}, \ldots, a_{pp})'$ by the vector of length p^2 formed by stacking the columns of $\mathbf{A}$. Also let $\mathbf{K}_p$ be the $(p^2 \times p^*)$ transition matrix (Browne, 1974, sec. 2; Nel, 1980, sec. 6) that relates the two vectors, such that vecs $(\mathbf{A}) = \mathbf{K}_p'$ vec $(\mathbf{A})$, and vec $(\mathbf{A}) = \mathbf{K}_p^{-\prime}$ vecs $(\mathbf{A})$, where $\mathbf{K}_p^- = (\mathbf{K}_p'\mathbf{K}_p)^{-1}\mathbf{K}_p'$ is the Moore-Penrose inverse of $\mathbf{K}_p$. Let $\mathbf{D}_k$ be the diagonal matrix defined by

$$\mathbf{D}_k = \mathbf{K}_p^-(\mathbf{K}_p^-)' = \operatorname{diag}(1, 2, 1, 2, 2, 1, \ldots, 1).$$

Since tr $(\mathbf{AB})$ = vec$'$ $(\mathbf{B})$ vec $(\mathbf{A}')$ (e.g., Graham, 1981, sec. 1.5), the derivative can be written as

$$\begin{aligned}\frac{\partial F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))}{\partial \gamma_i} &= -\text{vec}'\ (\mathbf{W}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{W}^{-1})\ \text{vec}\ (\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma})) \\ &= -\text{vecs}'\ (\mathbf{W}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{W}^{-1})\mathbf{D}_k\ \text{vecs}\ (\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma})) \\ &= \mathbf{b}_i'\ \text{vecs}(\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma})),\end{aligned}$$

where $\mathbf{b}_i = -\mathbf{D}_k$ vecs $(\mathbf{W}^{-1}\dot{\boldsymbol{\Sigma}}_i\mathbf{W}^{-1})$. Define the $(p^* \times q)$ matrix $\mathbf{B} = (\mathbf{b}_1, \ldots, \mathbf{b}_q)$. Then the gradient of (2) is

$$\frac{\partial F(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))}{\partial \boldsymbol{\gamma}} = \mathbf{B}'\ \text{vecs}\ (\boldsymbol{\Sigma}_0^* - \boldsymbol{\Sigma}(\boldsymbol{\gamma})) = \mathbf{B}'\tilde{\mathbf{e}}.$$

Clearly, if the value $\gamma = \gamma_0$ minimizes $F(\Sigma_0^*, \Sigma(\gamma))$ so that criterion (i) is satisfied, then the gradient must be null, that is,

$$\mathbf{B}'\tilde{\mathbf{e}} = \mathbf{0}|_{\gamma = \gamma_0}. \tag{5}$$

We proceed by first finding a vector, $\tilde{\mathbf{e}} = \text{vecs}\,(\tilde{\mathbf{E}})$, which satisfies (5). A coefficient, κ, is then found such that if $\mathbf{E} = \kappa\tilde{\mathbf{E}}$ and $\Sigma_0^* = \Sigma(\gamma_0) + \mathbf{E}$, then (ii) is satisfied with

$$M(\Sigma_0^*, \Sigma(\gamma_0)) = \delta \qquad \text{if } \mathbf{W} = \Sigma(\gamma_0) \tag{6a}$$

$$F(\Sigma_0^*, \Sigma(\gamma_0)) = \delta \qquad \text{otherwise.} \tag{6b}$$

Let $\mathbf{y}$ be a nonnull $(p^* \times 1)$ vector. The difference,

$$\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{B}\mathbf{v}, \tag{7}$$

is a vector satisfying (5) if we take $\mathbf{v} = (\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'\mathbf{y}$. Consequently, γ_0 is a stationary point of $F(\Sigma(\gamma_0) + \tilde{\mathbf{E}}, \Sigma(\gamma))$ when $\tilde{\mathbf{E}}$ is chosen in this manner. Criterion (i) requires the global minimizer of this function, not just a stationary point, however. To obtain the optimum value, we show that for

$$\Sigma_0^* = \Sigma(\gamma_0) + \kappa\tilde{\mathbf{E}},$$

γ_0 in fact is the global minimizer of $F(\Sigma_0^*, \Sigma(\gamma))$ as long as κ is not too large.

Proof. Let $\bar{\gamma}(\Sigma_0^*) = \arg\min_\gamma F(\Sigma_0^*, \Sigma(\gamma))$ be the overall minimizer. Then it can be shown that for any $\varepsilon > 0$, there exists a K such that $\|\bar{\gamma}(\Sigma_0^*) - \gamma_0\| < \varepsilon$ whenever $0 \leq \kappa \leq K$ (e.g., Kano, 1986, Theorem 1). We have shown that γ_0 is a solution to the system of equations

$$\frac{\partial F(\Sigma_0^*, \Sigma(\gamma))}{\partial \gamma} = \mathbf{0}, \tag{8}$$

and $\bar{\gamma}(\Sigma_0^*)$ is also a solution to these equations. We assume that $\Sigma(\gamma)$ is twice continuously differentiable in a neighborhood of $\gamma = \gamma_0$. It follows by the Implicit Function Theorem that for κ sufficiently small, there is a unique solution to (8) in a neighborhood $\|\gamma - \gamma_0\| < \varepsilon$. Consequently there exists a K such that $\gamma_0 = \bar{\gamma}(\Sigma_0^*)$ if $\kappa < K$. □

It remains to adjust $\tilde{\mathbf{e}}$ so that criterion (ii) holds, and here the two cases (6a) and (6b) must be distinguished. In the general case of (2) evaluated at $\gamma = \gamma_0$ so that (6b) holds, we have

$$F(\Sigma_0^*, \Sigma(\gamma_0)) = \tfrac{1}{2}\,\text{tr}\,[\mathbf{W}^{-1}(\Sigma_0^* - \Sigma(\gamma_0))]^2.$$

Defining $\mathbf{G} = \mathbf{W}^{-1}\tilde{\mathbf{E}}$, it is straightforward to solve

$$F(\Sigma_0^*, \Sigma(\gamma_0)) = \tfrac{1}{2}\,\text{tr}\,[\kappa\mathbf{G}]^2 = \delta$$

for κ to obtain

$$\kappa = [2\delta/(\text{tr}\,\mathbf{G}^2)]^{1/2}. \tag{9}$$

This completes the procedure for the general case.

In the particular case of maximum likelihood with argument $\gamma = \gamma_0$, the minimum value of the function has the form

$$M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) = \kappa \text{ tr } \mathbf{G} - \ln |\mathbf{I} + \kappa \mathbf{G}|,$$

where in this instance, $\mathbf{G} = \mathbf{W}^{-1}\tilde{\mathbf{E}} = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)^{-1}\tilde{\mathbf{E}}$. A solution for κ in closed form so that (6a) is satisfied is not apparent, but an iterative method can be applied. Let

$$\begin{aligned} t(\kappa) &= M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) - \delta \\ &= \kappa \text{ tr } \mathbf{G} - \ln |\mathbf{I} + \kappa \mathbf{G}| - \delta, \end{aligned}$$

with derivative

$$t'(\kappa) = \text{tr } \mathbf{G} - \text{tr } [(\mathbf{I} + \kappa \mathbf{G})^{-1}\mathbf{G}].$$

Obviously, $t(\kappa) = 0$ implies that $M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)) = \delta$. To solve the equation $t(\kappa) = 0$, the Newton-Raphson method is used. The update at the $(n+1)$-th step is

$$\kappa_{n+1} = \kappa_n - \frac{t(\kappa_n)}{t'(\kappa_n)}.$$

Using (9) with $\mathbf{G} = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)^{-1}\tilde{\mathbf{E}}$ gives an accurate starting value, κ_0, for the iterative sequence.

To summarize, the basic procedure requires computing the matrix $\mathbf{B}$, calculating the preliminary orthogonal component $\tilde{\mathbf{e}} = \mathbf{y} - \mathbf{B}\mathbf{v}$, and then rescaling to obtain the final version, $\mathbf{e} = \kappa\tilde{\mathbf{e}}$. For maximum likelihood, an iterative method is required to compute κ to satisfy the second criterion.

3. Notes on Implementation

Several points related to computation should be noted. To implement this algorithm, the partial derivatives $\dot{\boldsymbol{\Sigma}}_i$ are required in (4). These can be written explicitly for each model, but for the sake of generality, a numerical approximation might instead be preferred. For this approximation the finite forward difference method is

$$\dot{\boldsymbol{\Sigma}}_i = \frac{\boldsymbol{\Sigma}(\boldsymbol{\gamma} + \mathbf{u}_i h) - \boldsymbol{\Sigma}(\boldsymbol{\gamma})}{h},$$

where $\mathbf{u}_i$ is a vector with all elements equal to zero except for a single value of unity in the i-th position, and h is a small constant (i.e., $h = 10^{-8}$). When a computer program is written in this way, only the code describing the model $\boldsymbol{\Sigma}(\boldsymbol{\gamma})$ need be changed in each application.

In the least squares solution to (7), it is important to compute $\tilde{\mathbf{e}}$ accurately and efficiently, especially in problems where p or q are large. While forming the required cross-product matrices and solving normal equations is straightforward, it is not necessarily the most effective approach for this procedure. One may instead apply the QR factorization (e.g., Gill, Murray, & Wright, 1981, sec. 2.2.5.3) to the first q columns of the $(p^* \times q+1)$ matrix $(\mathbf{B}, \mathbf{y})$ to produce the transformed matrix $(\mathbf{B}^*, \mathbf{y}^*)$. This is followed by using the same component transformations, but applying them to $\mathbf{y}^*$ in reverse order, which then yields the vector $\tilde{\mathbf{e}}$. While this approach is less straightforward to implement than is an algorithm to solve normal equations, it is faster and produces very stable solutions. An excellent discussion of the computational details, along with its relative merits, is presented by Thisted (1988, sec. 3.1).

The preliminary residual vector $\tilde{\mathbf{e}}$ in (7) is the orthogonal complement of the projection of a vector $\mathbf{y}$ on $\mathbf{B}$. When maximum likelihood is used, $\mathbf{y}$ can be an arbitrary vector. With least squares, on the other hand, $\boldsymbol{\Sigma}_0^*$ is sometimes not positive definite

unless $\mathbf{y}$ is computed with some care. Consequently, we have employed the following method for calculating $\mathbf{y}$ which appears to be effective in eliminating the problem.

First, take $\mathbf{A}_1 = \mathbf{U}'\mathbf{U}$, where $\mathbf{U}(m \times p)$, $m > p$, consists of random uniform deviates, and rescale by

$$\mathbf{A}_2 = \text{Diag } (\mathbf{A}_1)^{-1/2}\mathbf{A}_1 \text{ Diag } (\mathbf{A}_1)^{-1/2},$$

so that $\mathbf{A}_2$ is at least positive semidefinite with unit diagonal elements. Next one must distinguish between those models that specify a measurement submodel for the observed variables and those that do not. Models that include a measurement submodel can be written in the form

$$\mathbf{\Sigma}(\boldsymbol{\gamma}, \boldsymbol{\Theta}) = \mathbf{\Omega}(\boldsymbol{\gamma}) + \mathbf{D}_\Theta, \tag{10}$$

where $\mathbf{\Omega}(\boldsymbol{\gamma})$ denotes the structural part of the model, and $\mathbf{D}_\Theta = \text{Diag } (\theta_1, \ldots, \theta_p)$ represents the unique variances of the p observed variables. Factor analysis models with unconstrained unique variances are the most obvious examples of structures with a measurement submodel. We write $\mathbf{\Sigma} = \mathbf{\Lambda\Phi\Lambda}' + \mathbf{D}_\psi$, where $\mathbf{D}_\psi$ is diagonal with positive values. This is an instance of (10) with $\mathbf{\Omega}(\boldsymbol{\gamma}) = \mathbf{\Lambda\Phi\Lambda}'$ and $\mathbf{D}_\Theta = \mathbf{D}_\psi$. When a model includes a measurement submodel of this kind, an additional scaling is applied, namely $\mathbf{D}_\Theta^{1/2}\mathbf{A}_2\mathbf{D}_\Theta^{1/2}$, so that $\text{Diag}(\mathbf{D}_\Theta^{1/2}\mathbf{A}_2\mathbf{D}_\Theta^{1/2}) = \mathbf{D}_\Theta$. Then in (7) let

$$\mathbf{y} = \text{vecs } (\mathbf{D}_\Theta^{1/2}\mathbf{A}_2\mathbf{D}_\Theta^{1/2}),$$

and proceed with the other steps as outlined above.

There are important classes of structures, for example patterned covariance or patterned correlation structures, which do not contain a measurement sub-model. For these cases, take $\mathbf{y} = \text{vecs } (\mathbf{A}_2)$ and proceed.

4. Computational Example

Jöreskog (1973, sec. 4) presented an artificial example of a seven variable covariance matrix with the structure given by

$$\mathbf{\Sigma}(\boldsymbol{\gamma}_0) = \begin{bmatrix} \mathbf{B}^{-1}(\mathbf{\Gamma\Phi\Gamma}' + \mathbf{\Psi})\mathbf{B}'^{-1} + \mathbf{D}_\varepsilon & \\ \mathbf{\Phi\Gamma}'\mathbf{B}'^{-1} & \mathbf{\Phi} + \mathbf{D}_\delta \end{bmatrix}. \tag{11}$$

The particular submatrices were

$$\mathbf{B} = \begin{bmatrix} 1 & -.8^* & -.4^* & 0 \\ 0 & 1 & 0 & -.3^* \\ 0 & 1 & 1 & -1 \\ -1 & 0 & 0 & 1 \end{bmatrix}; \quad \mathbf{\Gamma} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & .2^* \\ -1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}; \quad \mathbf{\Psi} = \begin{bmatrix} .2^* & & & \\ .1^* & .3^* & & \\ 0 & 0 & 0 & \\ 0 & 0 & 0 & 0 \end{bmatrix};$$

$$\mathbf{\Phi} = \begin{bmatrix} 1^* & & \\ .1^* & 2^* & \\ .2^* & .1^* & 3^* \end{bmatrix}; \qquad \mathbf{D}_\delta = \text{Diag } (.16^*, .36^*, .25^*);$$

$$\mathbf{D}_\varepsilon = \text{Diag } (.25^*, .36^*, .81^*, .25^*).$$

TABLE 1
Covariance Matrices for Model (11).

$\Sigma(\gamma_0)$

X_1	4.599						
X_2	2.481	2.069					
X_3	4.659	2.159	7.514				
X_4	6.449	3.731	7.409	10.799			
X_5	–0.692	–0.138	–1.454	–0.592	1.160		
X_6	2.100	1.250	2.750	4.100	0.100	2.360	
X_7	0.442	0.763	–0.421	0.542	0.200	0.100	3.250

$\Sigma^*_{0(M)}$, with $M(\Sigma^*_{0(M)}, \Sigma(\gamma_0)) = 0.25$ (maximum likelihood)

X_1	4.639						
X_2	2.568	2.195					
X_3	4.815	2.343	7.862				
X_4	6.545	3.814	7.675	10.593			
X_5	–0.461	–0.054	–1.385	–0.465	1.179		
X_6	2.323	1.407	2.997	4.182	0.205	2.471	
X_7	0.342	0.881	–0.154	0.251	0.246	0.227	3.250

$\Sigma^*_{0(L)}$, with $F(\Sigma^*_{0(L)}, \Sigma(\gamma_0)) = 0.25$ (least squares)

X_1	4.599						
X_2	2.627	2.069					
X_3	4.808	2.159	7.514				
X_4	6.298	3.738	7.412	10.795			
X_5	–0.585	0.015	–1.444	–0.612	1.160		
X_6	2.089	1.406	2.884	3.963	0.214	2.360	
X_7	0.495	0.877	–0.292	0.371	0.308	0.135	3.254

The starred coefficients are the population parameters, all other elements being considered known. The values δ_{33} and ε_{44} were specified to be equal. There are a total of $q = 19$ parameters in γ_0.

The value of $\Sigma(\gamma_0)$ implied by this structure is shown in the upper section of Table 1. When the above structure is applied to this matrix using maximum likelihood or least squares, the parameters are recovered exactly. The second section of the table contains the covariance matrix, $\Sigma^*_{0(M)}$, for which $M(\Sigma^*_{0(M)}, \Sigma(\gamma_0)) = 0.25$. When model (11) is fitted to this matrix by maximum likelihood, the parameter values are again as above, while the associated value of the discrepancy function is $M = 0.25$ (to four decimal places). The matrix **E** is the difference between the two tabled matrices. Similarly, the bottom section of the table contains the matrix, $\Sigma^*_{0(L)}$, for which $F(\Sigma^*_{0(L)}, \Sigma(\gamma_0)) =$

TABLE 2

Parameter Vectors when Applying Model (11) to $\Sigma^*_{0(M)}$ using Least Squares and to $\Sigma^*_{0(L)}$ using Maximum Likelihood.

	β_{12}	β_{13}	β_{24}	γ_{23}	φ_{11}	φ_{21}	φ_{22}	φ_{31}	φ_{32}
γ_0	−.80	−.40	−.30	.20	1.00	.10	2.00	.20	.10
γ_1	−.74	−.39	−.31	.34	1.08	.19	2.27	.30	.17
γ_2	−.76	−.44	−.29	.16	.88	.13	2.05	.15	.13

	φ_{33}	ψ_{11}	ψ_{21}	ψ_{22}	δ_{11}	δ_{22}	ϵ_{11}	ϵ_{22}	ϵ_{33}
γ_0	3.00	.20	.10	.30	.16	.36	.25	.36	.81
γ_1	1.93	.39	.09	.35	.08	.08	.02	.05	.42
γ_2	3.25	.10	.18	.19	.30	.43	.32	.64	1.30

	$\delta_{33} = \epsilon_{44}$
γ_0	.25
γ_1	1.31
γ_2	.00*

γ_1 is the parameter vector obtained from $\Sigma^*_{0(L)}$ using maximum likelihood as discrepancy function, while γ_2 is the parameter vector obtained from $\Sigma^*_{0(M)}$ using least squares as discrepancy function. *This parameter was on its lower bound.

0.25 when in (2) W = I. When the model is applied to this matrix with ordinary least squares, then the parameters are exactly recovered again.

It is known that if the model is applied to $\Sigma^*_{0(L)}$ using maximum likelihood, or alternatively, if it is applied to $\Sigma^*_{0(M)}$ using least squares, that the associated vectors of parameters are not the same, and in general neither are equal to γ_0 (Browne, 1969). Let these two vectors be denoted by γ_1 and γ_2, respectively. In Table 2 the two sets of parameters for these cases are shown. Some of the differences are rather marked. The values of the discrepancy functions are $M(\Sigma^*_{0(L)}, \Sigma(\gamma_0)) = 0.16$ for the matrix constructed under least squares but fit by maximum likelihood, and $F(\Sigma^*_{0(M)}, \Sigma(\gamma_0)) = 0.32$ for the matrix constructed under maximum likelihood but fit by least squares.

It was shown in section 2 that γ_0 minimizes $F(\Sigma^*_0, \Sigma(\gamma))$ if δ, or more precisely κ, is "small". If δ is "large", γ_0 may give a stationary point of $F(\Sigma^*_0, \Sigma(\gamma))$ that does not correspond to a minimum. To get some idea of the likelihood of this occurance, Σ^*_0 was

computed for all values of δ in steps of 0.25 between 0.0 and 16.0 (i.e., 0.0, 0.25, 0.50, . . .) using the same initial vector $\tilde{\mathbf{e}}$ in conjunction with maximum likelihood and the model in (11). Obviously, $\boldsymbol{\Sigma}_0^*$ must be positive definite, so δ cannot be too large. For example, at $\delta = 16.0$, $|\boldsymbol{\Sigma}_0^*| = 0.120 \times 10^{-5}$. Nonetheless, when $\boldsymbol{\Sigma}_0^*$ was used as the input matrix to fit (11), the correct parameters $\boldsymbol{\gamma}_0$ were always recovered, and the value of the discrepancy function at $\hat{\boldsymbol{\gamma}} = \boldsymbol{\gamma}_0$ was $M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma})) = \delta$. At larger values of δ the starting values for the iterative procedure had to be very close to the true value, however, otherwise local minima occurred. These were local minima because the true parameter vector was always associated with a smaller value of $M(\boldsymbol{\Sigma}_0^*, \boldsymbol{\Sigma}(\boldsymbol{\gamma}))$. It is possible that a value of $\boldsymbol{\gamma} \neq \boldsymbol{\gamma}_0$ exists that produces a smaller value of M for one or more $\boldsymbol{\Sigma}_0^*$, because local minima were sometimes a problem for cases with δ large. Based on these limited results, however, it appeared that $\boldsymbol{\gamma}_0$ was probably the global minimizer for $\boldsymbol{\Sigma}_0^*$ associated with each δ, which is reassuring. While it is unwarranted to generalize too much from one example, it appears that in practice δ may sometimes be relatively large and $\boldsymbol{\gamma}_0$ may still minimize the function.

5. An Application to the Factor Analysis of Ordered Categorical Variables

To illustrate how this method can be of value in empirical research, a small Monte Carlo experiment was designed to extend and partially replicate some results reported by Muthén and Kaplan (1985). They examined the performance of several factor analysis estimators with continuous, multivariate normal variables, but were especially interested in the estimators' behavior with non-normal, ordered categorical variables. To describe the data, let $\mathbf{z}^{*\prime} = (z_1^*, \ldots, z_p^*)$ denote p continuous, normally distributed variables with null mean vector. Let z_j be obtained from z_j^* by the transformation

$$z_j = \begin{cases} T & \text{if} \quad \tau_T \leq z_j^*, \\ \vdots & \qquad\qquad , \\ 1 & \text{if } \tau_1 \leq z_j^* < \tau_2, \\ 0 & \text{if} \quad z_j^* < \tau_1, \end{cases}$$

where τ_h, $h = 1, \ldots, T$, are thresholds common to all z_j^*. Muthén and Kaplan took $p = 4$ variables with $T = 4$. This gives observed variables with $T + 1 = 5$ categories. They described five conditions of categorization, defined according to different sets of thresholds. In this study, we used the first four of these sets:

Case 1:	−1.645	−0.643	0.643	1.645
Case 2:	−1.645	−1.036	−0.385	0.385
Case 3:	−1.881	−1.341	−0.772	0.050
Case 4:	−1.645	−1.282	−1.036	−0.674

An additional condition was based on the untransformed variables, $\mathbf{z} = \mathbf{z}^*$, hereafter denoted Case 0. Let $\mathbf{z}_{(g)}$ represent the variables obtained from Case g, $g = 0, 1, \ldots, 4$. Across these conditions, the transformed variables have increasingly larger coefficients of skewness and absolute value of kurtosis for both multivariate as well as univariate distributions (see their Table 1).

Because all components of $\mathbf{z}_{(g)}$ are related to those of $\mathbf{z}^*$ by a common transformation, the covariance structure of $\mathbf{z}_{(g)}$, $g \geq 1$, is the same as that for $\mathbf{z}_{(0)} = \mathbf{z}^*$. The structures have different parameter vectors, however. Let $\boldsymbol{\Sigma}(\boldsymbol{\gamma}_{(g)})$ denote the model

covariance matrix for $\mathbf{z}_{(g)}$, where $\boldsymbol{\gamma}_{(g)}$ is the associated parameter vector. Muthén and Kaplan studied a version of the factor analysis model for parallel tests,

$$\boldsymbol{\Sigma}(\boldsymbol{\gamma}) = \boldsymbol{\lambda}\boldsymbol{\lambda}' + \mathbf{D}_{\varphi}, \tag{12a}$$

in which

$$\boldsymbol{\lambda} = \mathbf{1}\lambda \quad \text{and} \quad \mathbf{D}_{\varphi} = \mathbf{I}\psi, \tag{12b}$$

with $\boldsymbol{\gamma}' = (\lambda, \psi)$. This structure was estimated as an unrestricted factor analysis model. Hence, $q = 8$ and df = 2. For $\mathbf{z}_{(0)}$, they specified $\boldsymbol{\gamma}_{(0)} = (0.7, 0.51)'$. A typical correlation between any pair of variables of $\mathbf{z}_{(0)}$ is therefore $\rho = \lambda^2 = 0.49$. Categorization introduces a bias in the model parameters and attenuates the resulting correlation (see footnote on their Table 1). Consequently, associated with the four sets of categorical variables, $\mathbf{z}_{(1)}, \ldots, \mathbf{z}_{(4)}$, were the parameter vectors (0.661, 0.563), (0.660, 0.564), (0.645, 0.584), (0.587. 0.655), respectively. These parameter vectors show the effects of increasing nonnormality in the categorical variables, inducing a downward bias in λ, and an upward bias in ψ.

Muthén and Kaplan (1985) found that maximum likelihood produced quite accurate estimates of the parameters in these conditions. They also found that the mean and variance of the computed maximum likelihood test statistic were approximately equal to the theoretical values for df = 2, except for the most extremely nonnormal condition, Case 4. Thus, the major conclusion of their study with respect to maximum likelihood was that the estimator gave accurate parameter estimates and a test statistic with approximately the correct distribution, except for the most nonnormal categorical data.

To extend these findings, maximum likelihood estimation of the model in (12) was studied with several population matrices constructed according to (1) assuming various degrees of lack of fit. In practicular, five values of $\boldsymbol{\Sigma}_0^*$ were generated for which δ in (6a) was equal to (0.0, 0.1, 0.2, 0.3, 0.4). This gave a total of twenty-five conditions (five values of δ by five categorization conditions). Samples of size $n = 1{,}000$ were generated, and each cell in the design was replicated 500 times. Mean parameter estimates were computed by averaging over all 500 replications and $p = 4$ variables. Means of the discrepancy function were computed across 500 replications. Results are summarized in Table 3.

In this experiment, the condition with $\delta = 0$ is the same as one part of the design studied by Muthén and Kaplan (1985). We review these results first. In terms of accuracy of parameter estimates, a pattern very similar to theirs was found. For all levels of $\mathbf{z}_{(g)}$, the mean parameter estimates followed the expected positive and negative bias in the parameters almost exactly. (The theoretical values are repeated in the bottom two lines of Table 3.) The mean value of the discrepancy function had a very slight increasing trend for more highly skewed conditions, a result that was quite similar to the analogous conditions summarized in Table 3 of Muthén and Kaplan for maximum likelihood.

In those conditions where the structure did not hold exactly, the mean parameter estimates were again essentially equal to the true values for all $\delta > 0$. Even with the most extreme case ($\delta = 0.40$), the parameter estimates were nearly identical to those for $\delta = 0$. In other words, maximum likelihood estimates appeared to be robust with respect to all levels of lack of fit. Differences that did occur only emerged for the various categorization conditions.

Next, consider results related to the discrepancy function. As anticipated for $\mathbf{z} = \mathbf{z}^*$ (Case 0), Mean(M) almost exactly equaled δ. This gives some assurance that the method works properly with continuous variables. A very different picture emerged

TABLE 3
Average Maximum Likelihood Parameter Estimates and Discrepancy Function Values for Model (12)

Population Discrepancy (δ)	Coefficient	Categorization Condition (Case g)				
		0	1	2	3	4
0.00	$\hat{\lambda}$[a]	0.699	0.662	0.660	0.646	0.588
	$\hat{\psi}$[a]	0.511	0.562	0.564	0.581	0.653
	M[b]	0.002	0.002	0.002	0.003	0.004
0.10	$\hat{\lambda}$	0.699	0.660	0.660	0.646	0.590
	$\hat{\psi}$	0.511	0.563	0.564	0.582	0.650
	M	0.103	0.065	0.072	0.064	0.045
0.20	$\hat{\lambda}$	0.701	0.661	0.661	0.646	0.588
	$\hat{\psi}$	0.509	0.562	0.562	0.582	0.651
	M	0.200	0.125	0.137	0.124	0.082
0.30	$\hat{\lambda}$	0.699	0.661	0.661	0.646	0.590
	$\hat{\psi}$	0.510	0.563	0.563	0.581	0.650
	M	0.302	0.182	0.198	0.178	0.118
0.40	$\hat{\lambda}$	0.699	0.661	0.660	0.646	0.590
	$\hat{\psi}$	0.512	0.562	0.563	0.581	0.649
	M	0.401	0.236	0.260	0.231	0.152
		Parameters[c]				
	λ_g	0.700	0.661	0.660	0.645	0.587
	ψ_g	0.510	0.563	0.564	0.584	0.655

[a]Means of $\hat{\lambda}$ and $\hat{\psi}$ were computed over 500 replications of 4 parameter estimates.
[b]Mean of M was computed over 500 replications.
[c]For Case g, $g \geq 1$, $\lambda_g = \sqrt{\rho_g}$ and $\psi_g = 1 - \lambda_g^2$, where ρ_g is the attenuated correlation due to categorization as reported in Muthén and Kaplan's Table 1.

from the four conditions with $\delta > 0$ in Cases 1 through 4, however. Compared to the conditions with $\delta = 0$, the average of the sample discrepancy function values *decreased* with increasing nonnormality. In other words, when the model did not hold and the distribution assumption was incorrect, the fit improved for cases with more extreme values of skewness and kurtosis. This finding is counterintuitive. For when $\delta = 0$, there was a slight deterioration in Mean(M) over levels of categorization, and across levels of $\delta > 0$ for fixed levels of g, Mean(M) worsened predictably. The effect of their joint

influence was not negative, as might have been anticipated, but rather produced an improvement in Mean(M) for the more extreme nonnormal cases.

These results demonstrate the potential utility of this method for constructing a covariance matrix with a prescribed lack of fit, and utilizing such a matrix in Monte Carlo experiments. It has led to the unexpected finding that there are situations where an incorrect assumption of normality leads to an unjustified impression that the model under consideration fits well. A more complete investigation of the behavior of estimators under lack of fit of the model, including additional simulation studies, is obviously needed, but this example illustrates that a rather different picture may emerge about an estimator's behavior when the assumption $\boldsymbol{\Sigma}_0 = \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0)$ is relaxed.

6. Discussion

A method has been presented for computing a population covariance matrix as the sum of a particular model plus a nonstochastic residual matrix, with the stipulation that the model holds with a prespecified lack of fit. It is felt that the procedure is promising for Monte Carlo research, and that sampling experiments based on populations where the model in (1) describes the covariance matrix are more realistic than those that assume a particular structure holds exactly. As the small experiment reported above illustrates, generating data in this fashion may yield results that are very different from those produced with the more common approach.

The present framework can be extended to apply to models with a structured mean vector as well as structured covariance matrix. This could be useful in several different settings, for example in the study of effects of assumption violation in certain analysis of variance problems (e.g., Huynh, 1978, sec. 5). The simplest way to incorporate this change is to parameterize both the covariance matrix and mean vector as components of an augmented matrix of scaled sums of squares and cross-products,

$$\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0) = \begin{bmatrix} 1 & \\ \boldsymbol{\mu}(\boldsymbol{\gamma}_0) & \boldsymbol{\Sigma}(\boldsymbol{\gamma}_0) + \boldsymbol{\mu}(\boldsymbol{\gamma}_0)\boldsymbol{\mu}(\boldsymbol{\gamma}_0)' \end{bmatrix},$$

where $\boldsymbol{\mu}(\boldsymbol{\gamma}_0)$ is the model for the vector of means. The order of $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0)$ is therefore $\tilde{p} \times \tilde{p}$, where $\tilde{p} = p + 1$, and the relationship in (1) becomes $\tilde{\boldsymbol{\Sigma}}_0^* = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0) + \tilde{\mathbf{E}}$. One modification must be made to the procedure as outlined above, however. The fixed value of unity in element (1, 1) of $\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0)$, $[\tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0)]_{11} = 1$, is simply a consequence of this form of parameterization, so the corresponding elements of $\tilde{\boldsymbol{\Sigma}}_0^*$ and $\tilde{\mathbf{E}}$ must be unity and zero, respectively. These conditions will be met if the first element of the vector $\mathbf{y} = (y_1, \ldots, y_{\tilde{p}})'$ in (7) is defined as

$$y_1 = (1 - c_{11})^{-1} \sum_{j=2}^{\tilde{p}} c_{1j} y_j,$$

where $(c_{11}, \ldots, c_{1\tilde{p}})$ is the first row of the matrix $\mathbf{C} = \mathbf{B}(\mathbf{B}'\mathbf{B})^{-1}\mathbf{B}'$. The remaining values $(y_2, \ldots, y_{\tilde{p}})$ are otherwise arbitrary as before. This choice for y_1 gives $[\tilde{\mathbf{E}}]_{11} = 0$, and $[\tilde{\boldsymbol{\Sigma}}_0^*]_{11} = \tilde{\boldsymbol{\Sigma}}(\boldsymbol{\gamma}_0)]_{11} = 1$.

References

Browne, M. W. (1969). Fitting the factor analysis model. *Psychometrika, 34*, 375–394.

Browne, M. W. (1974). Generalized least-squares estimators in the analysis of covariance structures. *South African Statistical Journal, 8*, 1–24.

Gill, P. E., Murray, W., & Wright, M. H. (1981). *Practical optimization*. San Diego, CA: Academic Press.

Graham, A. (1981). *Kronecker products and matrix calculus, with applications*. London: Ellis Horwood.

Hakstian, A. R., Rogers, W. T., & Cattell, R. B. (1982). The behavior of number-of-factors rules with simulated data. *Multivariate Behavioral Research, 17*, 193–219.

Huynh, H. (1978). Some approximate tests for repeated measurement designs. *Psychometrika, 43*, 161–175.

Jöreskog, K. G. (1973). A general method for estimating a linear structural equation system. In A. S. Goldberger & O. D. Duncan (Eds.), *Structural equation models in the social sciences* (pp. 85–112). New York: Seminar Press.

Kano, Y. (1986). Conditions for consistency of estimators in covariance structure models. *Journal of the Japanese Statistical Society, 16*, 75–80.

Laughlin, J. E. (1979). A Bayesian alternative to least squares and equal weighting coefficients in regression. *Psychometrika, 44*, 271–288.

Muthén, B., & Kaplan, D. (1985). A comparison of some methodologies for the factor analysis of non-normal Likert variables. *British Journal of Mathematical and Statistical Psychology, 38*, 171–189.

Nel, D. G. (1980). On matrix differentiation in statistics. *South African Statistics Journal, 14*, 137–193.

Shapiro, A., & Browne, M. W. (1988). On the asymptotic bias of estimators under parameter drift. *Statistics and Probability Letters, 7*, 221–224.

Thisted, R. A. (1988). *Elements of statistical computing*. New York: Chapman & Hall.

Tucker, L. R., Koopman, R. F., & Linn, R. L. (1969). Evaluation of factor analytic research procedures by means of simulated correlation matrices. *Psychometrika, 34*, 421–459.

Manuscript received 1/30/90
Final version received 8/28/91